U.S. DEPARTMENT OF EDUCATION

CRDC 2017-18 Technical Documentation

Release 1.0

September 2020

This technical documentation was produced under U.S. Department of Education Contract No. 47QTCA18D00JV with Applied Engineering Management Corporation. Stefanie McDonald serves as the contracting officer's representative. No official endorsement by the U.S. Department of Education of any product, commodity, service or enterprise mentioned in this publication is intended or should be inferred.

U.S. Department of Education

Betsy DeVos

Secretary of Education

CRDC

Stefanie McDonald

System Owner

September 2020

Document Control

DOCUMENT INFORMATION

Title:	CRDC 2017-18 Technical Documentation
Issue Date:	September 2020
Security Level:	Unclassified – For Official Use Only
Author(s):	Julia Bloom-Weltman Clover Quigley
Reviewer(s):	Sarah Meholick, Gabrielle Tapscott, Lauren Jetty, Marshal Fettro, Hoai Nguyen, Fred Johnson

DOCUMENT HISTORY

Version Number	Date	Summary of Change
1.0	September 2020	This is technical documentation for the public-use data file for the 2017-18 CRDC.

Table of Contents

1. Purpose

The purpose of the 2017-18 Technical Documentation is to provide data users with guidance and details pertaining to the 2017-18 Civil Rights Data Collection (CRDC). The following documentation provides general information on the CRDC. It describes different aspects of the 2017-18 data collection, including response rates, details for addressing data quality concerns, and privacy protection methodology. This documentation also includes an overview of post-collection outreach efforts and results.

2. About CRDC Data

The CRDC is a biennial survey that has been conducted by the U.S. Department of Education's Office for Civil Rights (OCR) since 1968. Since the 2011-12 collection, the CRDC has been a universal collection that collects data from all public local educational agencies (LEAs) and schools, including charter schools, alternative schools, schools serving students with disabilities, and long-term secure juvenile justice (JJ) facilities. [1]

The CRDC collects data on leading civil rights indicators related to access and barriers to educational opportunity at early childhood through 12th grade levels. The CRDC is also a longstanding and critical aspect of the overall enforcement and monitoring strategy used by OCR. It provides data that supports OCR's mission to ensure that recipients of the Department's federal financial assistance do not discriminate based on race, color, national origin, sex, or disability. OCR relies on the CRDC data it receives to:

- Investigate complaints alleging discrimination.
- Determine whether federal civil rights laws have been violated.
- Initiate proactive compliance reviews that focus on both acute and nationwide issues.
- Provide guidance and technical assistance to educational institutions and the communities they serve.

In addition, the CRDC is a valuable resource for other Department offices and federal agencies, policy makers and researchers, educators and school officials, parents and students, and all those who use data to better inform decisions concerning student equity and opportunity. For additional background information and FAQs regarding the CRDC, please visit: CRDC: Frequently Asked Questions.

2.1. Response Rates

Historically, the CRDC has achieved very high response rates. The 2017-18 CRDC included 17,637 LEAs and 97,763 schools (including JJ facilities and charter schools). Out of all respondents to the CRDC, 99.81% of LEAs, with 99.9% of schools, have certified data for the 2017-18 collection. This includes 17,604 LEAs and 97,632 schools. Less than 1% of LEAs did not certify data for the 2017-18 collection (34 LEAs). The list of LEAs that did not certify their data can be found in Appendix C.

2.2. Data Disaggregation

CRDC student count data are generally reported by race/ethnicity, sex, disability status, and English learner status. CRDC uses the 2007 race and ethnicity guidance published by the U.S. Department of Education. The guidance includes seven race and ethnicity categories (American Indian or Alaska Native,

[1] The 1976 CRDC collected LEA-level data from a universe of LEAs; a sample of those LEAs also provided school-level data. The 2000 CRDC collected from all LEAs and schools, as in the 2011-12 and subsequent collections.

Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, White, and two or more races). For more information on the Department's guidance regarding race and ethnicity categories, please visit New Race and Ethnicity Guidance for the Collection of Federal Education Data.[2]

3. Data Quality Suppression on the Public-Use Data File

Prior to its release, the public-use data file for the CRDC undergoes data quality suppression or reservation (DQ Reservation) which is standard methodology used to improve both the reliability and usability of data. The Department's Institute of Education Sciences' National Center for Education Statistics routinely uses data quality suppression methodology to improve the reliability and usability of its data. OCR's pilot process was implemented specifically to address potentially flawed or problematic data submissions. By targeting a subset of data from the most recent collection, the pilot was conducted on a small scale for the 2017-18 collection to evaluate the feasibility of implementation on additional modules in future collections.

For the 2017-18 collection, these methods were implemented on the following modules: Discipline (DISC)[3], Offenses (OFFN), Harassment or Bullying (HIBS)[4], and Restraint and Seclusion (RSTR). Each module was subject to data quality checks. In addition to these checks, some modules were verified using external data sources. Elements flagged[5] during these checks were then reserved.

Based on the results of DQ Reservation, some data in the public-use file have been replaced with a special reserve code.[6] Any data reserved due to data quality has a -11 as the reserve code in the public-use data file, meaning that -11 replaces the data submitted by the LEA and the original data are not available in the public files.

It is important to note that not all elements within each module were subject to such data quality methods. More details on the post-submission data quality checks used for DQ Reservation can be found in Appendix A.

4. Privacy Protection on the Public-Use Data File

In order to prevent the disclosure of identifying information, most data in the public-use file have been privacy protected by making small, random adjustments to the data. This process, also called perturbing, used a low-frequency perturbation routine. The methodology applied to the data added or subtracted one case to blur the data and used random data swapping. The routine protected true zeros, except for outcome data (e.g., Algebra passing), for which zeros are included in the perturbation routine. The routine was applied to all student count data elements. It is important to note that the perturbation

[2] Guidance surrounding race and ethnicity has not changed from the previous collection.
[3] Data quality suppression for the DISC module was limited to questions involving corporal punishment.
[4] While there is a Harassment and Bullying (HIBD) module at the LEA level, the Harassment or Bullying (HIBS) module used in DQ Reservation is at the school level.
[5] The word "flag" indicates whether a LEA is identified as having a particular data quality issue.
[6] Reserve codes are negative values within the data set that are not an actual count. They represent variables that do not have reported values, or values that have been suppressed due to data quality. A list of reserve codes and their definitions can be found in the CRDC User Manual.

routine rolls up to the specified change thresholds at the state and national levels, but there are no threshold for roll ups at the LEA level.

There were several bounds set for these privacy protections. Perturbed data counts for outcome data were bounded to not exceed the number of students who took a class. Perturbations of related data elements (i.e., student counts and instances) were bounded so that perturbed data counts for the number of students do not exceed the number of instances. Finally, the margin of difference between perturbed and unperturbed counts varied by the size of the data sample being perturbed. The perturbation routine was re-run until it fell within the acceptable ranges.

5. Post-Collection Outreach to Correct Data Quality Errors

As part of a comprehensive data quality review process, OCR identified priority data issues based on data anomalies. These anomalies were identified during and after the CRDC submission period, ending on June 21, 2019, through a series of specific and general data quality checks. Outreach was then conducted for LEAs identified through these data quality checks. In response to these data quality checks, LEAs were given an opportunity to correct and recertify their data in the summer of 2019; this is otherwise known as a data correction. A data correction is revised data for a data element (e.g., enrollment numbers for female students).

In total, outreach was conducted on almost 90% of LEAs who submitted data for the 2017-18 collection. This was a marked increase from the 2015-16 collection when outreach was targeted to a smaller subset of LEAs. The expanded effort was endeavored by OCR to improve data quality.

5.1. Identifying Data Anomalies

The data anomalies used to inform outreach were identified through a series of specific and general data quality checks. While the specific data quality checks are specific to certain modules, the general checks — which were piloted for the 2017-18 collection — are meant to cover the majority of data elements in the CRDC. They are designed to uncover extreme values in terms of: 1) potentially anomalous data within a single data element, 2) a change from the previous collection, or 3) anomalous data among all responses in a module.[7]

In total there were 36 specific data quality checks included in outreach, some of which were broken into sub-parts. In addition to these specific checks, three general checks, described above, were used. Details and definitions for the specific and general data quality checks included in outreach can be found in Appendix B.

In response to the results of the data quality checks, OCR reached out to LEAs immediately after the collection closed to remedy potential data quality issues. OCR also reached out to all LEAs to verify reports of all zeros within the Restraint and Seclusion module.

[7] Extreme values, considered potential outliers, were determined using a threshold of four standard deviations based on univariate distributions and based on changes between collections. We also flagged outliers using a multivariate clustering technique.

Thirty-seven percent of LEAs included in outreach either amended their data, submitted a data note, or both. If the LEA determined that the data identified for review needed correction, LEAs were able to make corrections within the data submission tool. Of the LEAs that submitted explanations, a large portion indicated that their data were correct-as-reported.[8]

5.2. Explanations for Data Quality Errors

As stated earlier, after the CRDC submission period closed LEAs flagged for a particular data quality check had the opportunity to submit an explanation and/or correct their data. If an LEA could not make changes to the data, or believed that no changes were needed, the LEA was asked to submit an explanation justifying their data submission.[9]

5.3. LEAs Flagged By Data Quality Checks in the Public-Use Data File

The accompanying data tables, located in Appendix B, contain 1) definitions for the data quality checks included in outreach, 2) LEAs flagged with these checks in the current public-use data file,[10] and 3) whether LEAs submitted correct-as-reported explanations for identified issues during outreach. The purpose of these tables is to provide detail on the data quality issues identified, and in some cases corrected, during post-collection. The LEA-level detail also provides information about those LEAs that still have data quality issues in the public-use file. For those LEAs who received outreach on data quality checks, we also provide where LEAs stated their data were correct-as-reported despite the issue identified.

In summary, 12,218 out of the 17,604 LEAs that certified their data were flagged by at least one data quality check in the public-use file, which can be seen in the summary sheet in Appendix B. It is important to note that these accompanying data tables include those LEAs flagged by data quality checks in the public-use datafile and do not include all LEAs included in outreach originally. On average, LEAs were flagged with two data quality checks when analyses were conducted on the public-use data file.

6. Additional Corrected Data Errors

In addition to data corrections made as a result of outreach in the summer of 2019, OCR included an extended correction period for data quality issues in the fall of 2019 (ending December 15, 2019). The public-use file does not include data corrections made after December 2019.

6.1. Force Certification

It is important to note that as part of the data corrections process after the collection close, there were a large number of force certifications in post-collection. Forced certification indicates instances where a submission system business rule error was unresolved by the data submitter, and the Partner Support Center manually reviewed the data submission and certified the data on the data submitter's behalf.

[8] The term "correct-as-reported" indicates that a LEA submitted an accompanying explanation that verifies their data were correct as initially reported in response to a data quality check. For example, an LEA flagged with inconsistencies in student enrollment may indicate their responses are correct-as-reported because they are a high mobility school.

[9] Only LEAs that provided documentation for data notes in the correct format were processed.

[10] Outreach was conducted on an earlier extract of data from the submission system.

Normally, data that are force certified require the submission of an action plan. Action plans are required in cases where a school or LEA is unable to provide data for the CRDC, or when data is force certified. An action plan is meant to ensure that accurate data are submitted for future collections.

The large number of force certifications in the 2017-18 collection was due to the changing of data in the Restraint and Seclusion module from zeros to nulls, meaning that the data were not collected. As mentioned earlier, a focus of the post-collection data quality outreach was on ensuring the distinction between zeroes and nulls and reporting correctly, particularly for the Restraint and Seclusion module. For this collection, data force certified for this reason during the post-collection work were not required by OCR to have an action plan.

7. Other Data Anomalies

There were instances when the data submission system did not function as expected, or where other data anomalies occurred in the public-use file. A summary of those issues follows.

7.1. New York City Public Schools Action Plan

For the 2017-18 school year, an action plan was issued for the New York City Public Schools due to a large amount of missing data. Because of the size of the LEA, the electronic action plan timed out before the submission system could populate the missing data. Therefore, the district submitted a paper action plan and was force certified. Instead of receiving a reserve code indicating an action plan, this district was incorrectly assigned a reserve code indicating force certification. Thus, the action plan reserve code was not accurately displayed in New York City's data. Table 2 lists the CRDC questions and elements affected by this anomaly.

Table 1. Elements Affected By New York City Public Schools Action Plan

Question	Title	Elements Affected
SECR-1	Security Staff	All data elements
ARRS-1b	Discipline of Students Without Disabilities – School-Related Arrest	All data elements
ARRS-2b	Discipline of Students With Disabilities – School-Related Arrest	All data elements
RSRT-1a	Non-IDEA Students Subjected to Restraint or Seclusion – Mechanical Restraint	All data elements
RSTR-2a	IDEA Students Subjected to Restraint or Seclusion – Mechanical Restraint	All data elements
RSTR-3	Instances of Restraint or Seclusion	Data elements RSTR-3.1 - RSTR3.3
OFFN-1	Offenses – Number of Incidents	Data elements OFFN-1.1 - OFFN-1.11

7.2. Skip Logic

For the 2017-18 collection, there were instances where the skip logic[11] in the Student Discipline - Referrals to Law Enforcement & School Related Arrests (ARRS) module did not function as expected. In this module, the skip logic functionality failed to update related data fields. This functionality failure allowed data elements to require data entry, when according to their skip flag those elements should

[11] Skip logic is a submission system functionality that redirects respondents to different points in the survey based on answers to specific questions.

have been skipped. Conversely, there were also elements that required data entry according to their skip flag, but were skipped due to this skip logic failure. The failure in the skip logic occurred due to intense system load, which may cause the system to slow down and not update skip logic flags effectively. Any schools involved in this skip logic failure will have a "-3" reserve code listed for the module.

7.3. Misalignment Across Student Counts and Instances of Out-Of-School Suspension (OSS)

A subset of schools in the 2017-18 public-use data file for the CRDC report fewer instances of out-of-school suspension (OSS)[12] than students identified as receiving OSS[13] in the Student Discipline (DISC) module. This is logically inconsistent with OCR guidance. The total count of instances of OSS should always be greater than or equal to the total count of students identified as receiving OSS, because each student receiving OSS should only be reported once, regardless of how many times the student was suspended.

The misalignment across student counts and instances of OSS is a data quality issue that occurred as a result of incorrect data submissions (despite outreach conducted to LEAs regarding this issue). This relationship will be preserved in future data collections by including a submission system business rule that checks that the data was inputted correctly and by maintaining the relationship as part of the perturbation requirements.

8. Appendices

8.1. Appendix A: Data Quality Reservation Table for Public-Use Files

Below we provide a description of the DQ Reservation methods applied to the 2017-18 public-use data.

Module	Check Type	Description	Information Not Shown in the Public Release
Restraint & Seclusion (RSTR)	General	**Year-to-Year**: Flag when absolute difference and percentage difference both exceed four standard deviations. [Exceptions: 1) Data flagged as univariate outliers in 2017-18 are excluded. 2) Data where there were zeros in 2015-16 and values in 2017-18 were excluded.]	Any individually flagged data element, including related calculated totals
RSTR	General	**Multivariate:** Flag when responses do not cluster with other schools in ways that appear erroneous based on a multivariate analysis and a manual secondary review.	Entire module

[12] Instances of OSS are reported in question DISC-11 of the CRDC submission system.
[13] Students identified as receiving OSS are reported in questions DISC-7c and DISC-9c of the CRDC submission system.

Module	Check Type	Description	Information Not Shown in the Public Release
RSTR	New	**External Data Validation- News:** School or LEA data reports zero incidents of restraint and seclusion in 2017-18 data but media reports incidents of restraint and seclusion during the same time frame.	Entire module
RSTR	New	**External Data Validation- RSTR Data Quality Review:** For those districts that were a part of OCR's 2015-16 Data Quality Review, flag LEAs that submitted zeros or no data for all elements in the module for both 2015-16 and 2017-18.	Entire module
RSTR	Specific	**Check 57:** Student count by race-ethnicity/sex should not be identical across mechanical restraint, physical restraint, and seclusion for (A) IDEA and (B) non-IDEA) *Threshold:* greater than 10 students for any duplicate count.	If any race/ethnicity flagged, all flagged RSTR student count data elements for IDEA and non-IDEA
RSTR	Specific	**Check 58:** Non-zero instances of mechanical restraint, physical restraint, and seclusion (A) IDEA and (B) non-IDEA) should not be identical. *Threshold:* greater than 10 instances.	If any race/ethnicity flagged, all flagged RSTR student count data elements for IDEA and non-IDEA
RSTR	Specific	**Check 59:** Reported student counts subjected to mechanical restraint, physical restraint, and seclusion by sex and race/ethnicity should not be larger than student enrollment by sex and race/ethnicity (A) IDEA and non-(B) IDEA *Threshold:* if number of students subjected to mechanical restraint, physical restraint or seclusion is greater by a magnitude of 10 or more students compared to the total enrollment in a school.	If any race/ethnicity flagged, all flagged RSTR student count data elements for IDEA and non-IDEA

Module	Check Type	Description	Information Not Shown in the Public Release
RSTR	New-Specific[14]	**Check 74:** Counts of total students should not exceed number of instances by the following groupings: 1. Mechanical restraint- without disability 2. Mechanical restraint- IDEA 3. Mechanical restraint- 504 4. Physical restraint- without disability 5. Physical restraint- IDEA 6. Physical restraint- 504 7. Seclusion- without disability 8. Seclusion- IDEA 9. Seclusion- 504	All flagged data elements Checked separately for #1-9
Offenses (OFFN)	General	**Year-to-Year**: Flag when absolute difference and percentage difference both exceed four standard deviations. [Exceptions: 1) Data flagged as univariate outliers in 2017-18 are excluded. 2) Data where there were zeros in 2015-16 and values in 2017-18 were excluded.]	Any individually flagged data element
OFFN	General	**Multivariate**: Flag when responses do not cluster with other schools in ways that appear erroneous based on a multivariate analysis and a manual secondary review.	Entire module
OFFN	New	**External Data Validation:** School or LEAs tracked in shooting database "Everytown" as having previous incidents/reports of firearm possession or use in 2017-18, but the school or LEA did not include incidents/reports in their data submission.	All flagged data elements
OFFN	Specific	**Check 66**: The total number of offenses reported by schools within large LEAs (enrollment greater than or equal to 25,000 students) should not be zero.	Entire module
OFFN	Specific	**Check 67**: The sum of reported offenses should not be greater than or equal to the total student enrollment. *Threshold*: Absolute difference over 10 and the percentage difference was over 200%.	Entire module
OFFN	New-Specific	**Check 75:** Counts of offenses within the following data elements should not be identical: 1) all data elements in question OFFN-1, 2) threat incidents, 3) attack incidents, 4) robbery incidents. *Threshold:* 10.	1) Entire module; 2) All threat incidents

[14] "New-Specific" checks are specific checks that were developed for data quality suppression. These checks will also be used in future collections.

Module	Check Type	Description	Information Not Shown in the Public Release
			3) All attack incidents 4) All robbery incidents
OFFN	New-Specific	**Check 76**: Counts of offenses should not exceed expected range. A. Rape>15 B. Robbery with a firearm>29 C. Possession with a firearm>75 (4 standard deviations)	Any individually flagged data element
Harassment or Bullying (HIBS)	General	**Year-To-Year:** Flag when absolute difference and percentage difference both exceed four standard deviations. [Exceptions: 1) Data flagged as univariate outliers in 2017-18 are excluded. 2) Data where there were zeros in 2015-16 and values in 2017-18 were excluded.]	Any individually flagged data element, including related calculated totals
HIBS	General	**Multivariate:** Flag when responses do not cluster with other schools in ways that appear erroneous based on a multivariate analysis and a manual secondary review.	Entire module
HIBS	New	**External Data Validation- News:** School or LEA reports zero incidents of bullying or harassment when news articles report previous incidents/patterns of bullying or harassment.	Entire module
HIBS	Specific	**Check 72**: Allegations/reports/disciplined students should not be greater than enrollment. *Threshold*: Difference between instances and enrollment > 10.	All flagged data elements
DISC	New-Specific	**Rule 77:** Instances of corporal punishment are less than number of students who received corporal punishment: a) all preschool children, b) preschool children with disabilities (IDEA), c) K-12 students without disabilities, d) K-12 students with disabilities.	All flagged data elements

8.2. Appendix B. LEAs Identified with Post-Submission Data Quality Checks

The linked data tables contain (a) LEAs flagged with post-submission data quality checks included in outreach using the public-use data file, (b) definitions of these data quality checks, and (c) whether LEAs submitted correct-as-reported explanations for identified checks. The “Read Me” tab outlines guidance for how to navigate through the tables.

8.3. Appendix C. LEAs That Did Not Certify

The following LEAs did not certify their data by the close of the CRDC.

LEA ID	LEA Name	State Abbreviation
0100300	Barbour County	AL
0400390	Camelback Education Inc	AZ
0400613	Educational Impact Inc.	AZ
0625470	Montebello Unified	CA
0699744	Foothill Leadership Academy	CA
06CC293	Collegiate Charter High School of Los Angeles	CA
1100079	National Collegiate Preparatory PCHS	DC
1700003	Blue Ridge CUSD 18	IL
1704380	Ashton-Franklin Center CUSD 275	IL
1810110	M S D Shakamak Schools	IN
2600080	El-Hajj Malik El-Shabazz Academy	MI
2630120	Romulus Community Schools	MI
2700153	Harvest Preparatory School	MN
2700335	Hennepin Elementary School	MN
3303292	PACE Career Academy Charter School	NH
3400752	The Kingdom Charter School of Leadership	NJ
3405370	Franklin Lakes School District	NJ
3500116	Alma D'Arte Charter	NM
3501530	Las Vegas City Public Schools	NM
3502160	Questa Independent Schools	NM
3502490	Springer Municipal Schools	NM
3600001	Dolgeville Central School District	NY
3600163	International Leadership Charter School	NY
3601133	Charter High School for Law and Social Justice	NY
3901473	Hope Academy for Autism	OH
4000794	Langston Hughes Academy for Arts and Technology	OK
4005160	Bowring	OK
4009000	Crescent	OK
4015450	Indianola	OK
4016410	Keyes	OK
4026640	South Coffeyville	OK
4030048	Woodland	OK
4105640	Glendale SD 77	OR
4700147	Achievement School District	TN